PHYSICIAN INFORMATION
for Breast Companion® CADx Software System

INDICATIONS FOR USE

Breast Companion® is a computer-aided system (CADx) intended for improving the ACR BI-RADS® assessment of ultrasound images of lesions of the female breast as part of the diagnostic workup.

CONTRAINDICATIONS

There are no contraindications for the use of this device.

WARNINGS

Warnings related to the safe and proper use of the Breast Companion® software are displayed where applicable in the software user interface.

PRECAUTIONS

Precautions related to the safe and proper use of the Breast Companion® software are displayed where applicable in the software user interface.

The Breast Companion® software should be used only by ACR BI-RADS® qualified and trained personnel. ACR BI-RADS® Clinical, Lexicon and Classification guidance should be followed when reviewing images using Breast Companion®.

User Manual guidance on software operation should be followed.

Imaging studies with the following characteristics are not recommended for use with the Breast Companion's Computer-aided Lesion Assessment (BC CLA) scoring function:

- Studies with BI-RADS categories of "No findings" , "Incomplete Data" or "Known Cancer" for a selected image file
- Fewer than 2 views of the mass available
- Graphic overlays, calipers, foreign objects or markers overlaying the lesion in question image
- Excessive artifacts or only color Doppler images present.

Inaccurate or tentative lesion definition (including outlined lesion border contour) by the user may lead to inaccurate CLA scoring.

Almen Laboratories, Inc.
1672 Gil Way, Vista CA 92084
(760) 806-0040(voice), (760) 806-7577(fax)
e-mail: info@almenlabs.com http://www.almenlabs.com

DEVICE DESCRIPTION

Breast Companion® ("BC", "BC CADx") is identical to BI-RADS Companion™ (K072258) with an added optional scoring function (Computer-aided Lesion Assessment, or "CLA", or "BC CLA"). It is an adjunctive tool to be used in support of radiologists' readings and is intended to be used in the diagnostic breast work-up process. This is an important distinction from computer-aided detection devices for screening. CLA is intended to be used by radiologist as decision stratification tool.
The two components are related as indicated in Figure 1 below.

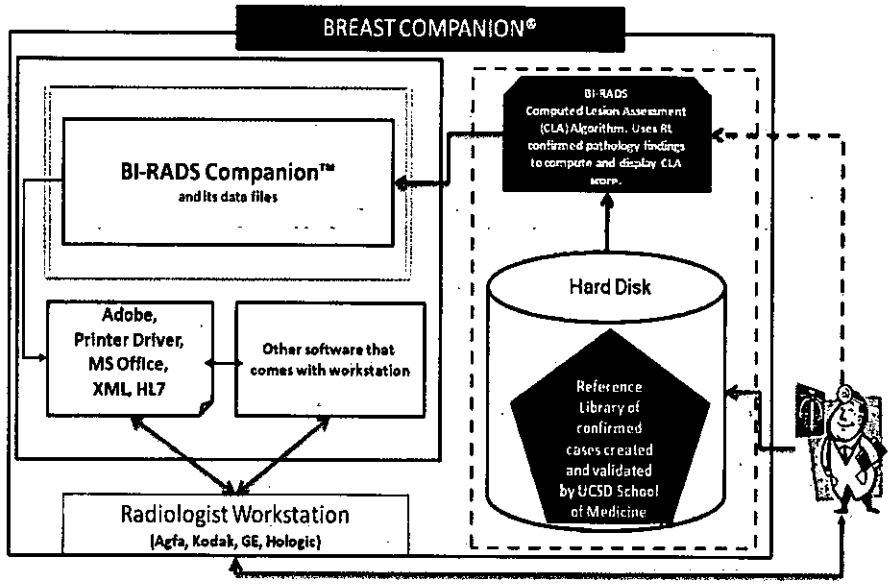

Figure 1. Breast Companion® upgrade to a BI-RADS Companion™ (K072258) installation in a typical user scenario

The BC device includes a report-generating function that is fully compliant with the American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS®). BC CLA is optionally used by a Radiologist who is completing an ACR BI-RADS® assessment report of a lesion that was detected or identified in a previous procedure using a different modality such as screening mammography or physical examination, or patient discovery during self-examination. This ACR BI-RADS® report is a record of the sonographic clinical assessment and incorporates recommendations by the radiologist for further action within the overall diagnostic process.
The BC CLA function is designed to work as an "upgrade" feature on already installed BI-RADS Companion™ CAD software (Class II, K072258). BI-RADS Companion™ may be purchased with the BC CLA function activated – that configuration is called Breast Companion®.

Breast Companion® functionality working as an enhancement of BI-RADS Companion™ computer-aided system with BC CLA function available to the user and can be summarized as follows:

When invoked by the user, the BC CLA function receives calculated numeric imaging parameters of the lesion previously defined (outlined) by the user and already stored in computer memory

↓

Based on calculated numeric imaging parameters of the lesion in question, Breast Companion® compares and retrieves the 7 most similar lesions from the Reference Library of cases with confirmed findings

↓

BC CLA function computes the CLA score based on confirmed pathology indicator of the retrieved most similar lesions and displays the CLA score using Breast Companion® menu system

SUMMARY OF CLINICAL STUDIES

There were multiple research and then clinical validation studies conducted to evaluate the safety and effectiveness of the Breast Companion® CADx Software System. The description below is a brief summary of those studies. The studies are published in detail in more than 20 scientific publications (including multiple peer reviewed) and results were reported were presented at SPIE 2001-2004, RSNA 2003-2011, AAPM 2003, 2009-2011, AIUM 2008-2011, AI 26-30. R&D stages of the studies were supported in part by multiple NIH/NCI, SBIR and 3 private research grants.

Contributors & Collaborators

Two luminary sites directly participated in clinical validation of the software – Thomas Jefferson University Hospital (including the Breast Imaging Center of TJU) in Philadelphia, PA and the UCSD School of Medicine (lead site) in San Diego, CA. All clinical data, clinical validation and clinical analyses of the results were developed, collected, processed and analyzed in those institutions under their respective IRB protocols. The readers represented a variety of experiences in sub-specialty from minimal to extensive. The readers assigned to the Validation Study were randomly selected by the validation sites.

Validation Study Design

The following classical methods and endpoints were used:

- MRMC study design
- DBM method in estimation of validation Sample Size
- ROC AUC as primary endpoint,
- Specificity and Sensitivity as secondary endpoints.

In the Breast Companion® study, ROC curves for BI-RADS scores assigned to a cohort of breast ultrasound imaging studies by multiple sub-specialty radiologists were compared "without" and "with" the use of the BC CLA.

Validation Case File Selection and clinical data HIPAA compliance and security

Potential validation case files were initially identified by the clinical validation group from the catchment population by searching chronologically through their RIS database. This PACS-based digital database consists of files from studies (Joint Commission inspections for "standard of good practice" and HIPAA Regulations applied) performed on patients referred to their respective breast imaging clinics for diagnostic workup or biopsy. All case files were obtained retrospectively from PACS. Only those cases with a corresponding confirmed findings (biopsy result or a documented two-year negative follow up) were "enrolled" in the study and the cases were retrieved in consecutive chronological order. No patients or cases were produced specifically for this study or per request from Almen Laboratories. Therefore the catchment represents a true population pool that includes all ethnic and racial groups, all age groups, and all tests completed on ultrasound stations in clinical use across all participating localities.

All clinical data used in validation studies was IRB approved and had verified confirmed findings. A minimum of two images for each mass were required for the case to be ruled eligible for inclusion in the study. Minimal US image artifacts were allowed, but each view had to be free of graphic overlays, calipers and other markers, including color Doppler. Studies with no lesion present or known cancers were excluded. The eligible cases with all eligible available images and views were then entered into the Validation Database for analysis "without" and "with" the BC CLA function.

The final validation database consisted of 596 archived image cases. The final case mix was as follows:

Lesion type	N
Simple Cysts	165 (27.7%)
Complicated Cysts	58 (9.7%)
Solid Benign	240 (40.3%)
Malignant	133 (22.3%)
Total	596 (100%)

Table 1. Case Mix, All Cases

Size Distribution

The analysis included a histogram distribution of lesion size as well as complete descriptive statistics. The range was 2.0 to 98 mm, mean 12.8, median 10.0, standard deviation 9.3. The 95% reference limit was 3.0 mm (95% CI 2.0-4.0 mm) for the lower limit and 34.1 mm (95% CI 31.0-52.0 mm) for the upper limit.

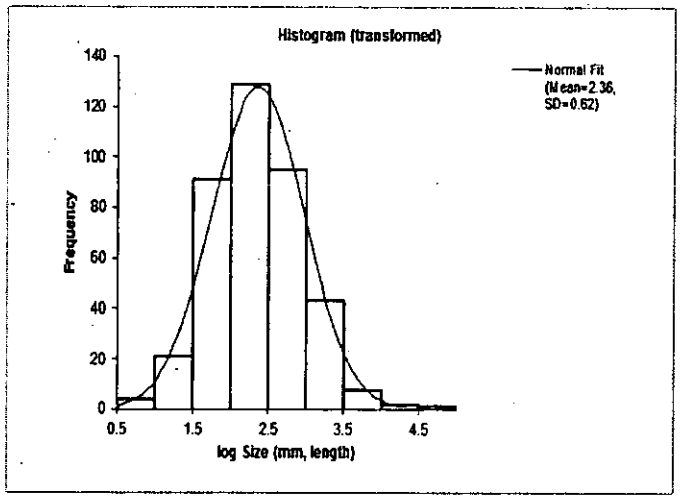

Fig. 2. Lesion size distribution in Validation dataset

U/S Scanner Distribution

The ultrasound imaging studies used in the Validation Study were performed on a variety of makes and models of scanners. The studies used in all clinical studies are based on "standards of good practice" in breast ultrasonography, ACR BIRADS Guidance, and standards of practice in an MQSA accredited facility[1].

Truth

Truth was established for the 596 cases retrieved for the final validation dataset by confirming biopsies and pathology reports in 390 cases (65.4%), while the remaining 206 cases (34.6%) that did not have biopsies had benign clinical follow-up of at least 2 years. Of the 206 cases that were recommended for follow-up and did not require biopsy, the characterizations by the radiologist were as follows:

Lesion type	N
Simple Cysts	165 (80.1%)
Complicated Cysts	21 (10.1%)
Solid Benign	20 (9.7%)
Total	206

Table 2. Case Mix, Follow-up Cases

Patient assessment and demographic data

No human subjects were recruited for the study because only previously archived pre-existing breast US files were eligible for enrollment. Potential cases were identified by the clinical validation group as previously described. While subject cases were not specifically screened for

[1] *Helmut Madjar, Ellen B. Mendelson, The Practice of Breast Ultrasound: Techniques, Findings, Differential Diagnosis, Thieme Verlag, Germany, 2008*

race, ethnic background or health status, the class of subjects included only women who had diagnostic sonograms and other related procedures at the study institution (i.e. biopsy, traditional breast ultrasound, MRI, etc.). The study did not specifically include or exclude any ethnic groups. No case files were excluded from the study after enrollment. Because the study used pre-existing cases with patient identifiers removed as a condition of the study (HIPAA Regulations, 2005), a complete demographic profile of the study population was not analyzed.

Validation Studies period

The Validation Study period was for approximately 3 years, from late 2006 – late 2009. The initial IRB approval for the validation dataset was received in late 2006 and the study began shortly thereafter. Annual IRB reports were made to document study progress, and the IRB granted continued approval through the study period.

Device Standalone Performance Data Analysis

Standalone performance testing was designed to demonstrate how the BC CLA scoring algorithm performed on the lesions that are outlined (segmented) by the radiologists compared to the confirmed truth of the cases. ROC-AUC, Sensitivity and Specificity were the endpoints of this evaluation. The "standalone" computing performance evaluation provided safety data for the BC CLA function. Specifically the question to be answered was, "*Does the BC CLA function, which is subject of this PMA, produce computed-lesion assessments score corresponding to the appropriate BI-RADS Category for a lesion in the absence of physician interaction at the score computing phase?*"

The diagnostic Sensitivity of the BC is defined as the conditional probability that a person having a disease will be correctly identified: TP/(TP+FN). The diagnostic Specificity of the BC is defined as the conditional probability that a person not having a disease will be correctly identified: TN/(TN+FP). True Positive, True Negative, False Positive, and False Negative definitions are used in accordance with traditional references.

The Validation Case Files cohort was used to produce BC CLA scores that were consequently analyzed with ROC analysis. The 596 lesions assessed by the radiologist were scored using the Reference Library cases in batch processing using the radiologist's segmentation but without other radiologist interference in the CLA scoring. As a result, 596 BC CLA scores were produced and compared to the confirmed truth of each case.

This "standalone" computing performance evaluation was performed six-to nine months prior to the "without" reading under the supervision of a single lead expert radiologist using the 596 validation cases and the Reference Library, as in the proposed device. The radiologist did not participate in the BC CLA scoring that was done and recorded automatically as a direct export from the BC software. The radiologist's supervision was limited to border definition of the lesion-in-question that was confirmed or corrected by the radiologist. Since BC CLA is intended to be used with the radiologist in the loop and under radiologist supervision this was the correct

approach to evaluating "standalone" computing performance. When the lesion border was confirmed by the radiologist, the BC CLA scoring function was then used in automatic batch mode (no presence of any clinical personnel) and the BC CLA score produced was automatically recorded for each case.

Table 3 below presents the ROC AUC analysis of the standalone performance of the BC CLA scoring function.

Test	Area (A_z) (fitted)	Area (A_z) (empirical)	95% CI	SE	Z	*p*
BC CLA	98.6%	98.2%	0.97 to 0.99	0.006	82.81	<0.0001

H_0: Area ≤ 0.5. H_1: Area > 0.5.

Table 3. ROC AUC (A_z) analysis of standalone performance (BC CLA)

Table 4 below presents the Sensitivity and Specificity data for the BC CLA.

BC CLA Score (Positive test > cutoff)	TP rate (Sensitivity)	95% CI		TN rate (Specificity)	95% CI	
Cutoff 2.5	0.985	0.947	to 0.998	0.890	0.858	to 0.917
Cutoff 3.0	0.955	0.904	to 0.983	0.946	0.921	to 0.965
Cutoff 3.5	0.932	0.875	to 0.969	0.961	0.939	to 0.977
Cutoff 4.0	0.865	0.795	to 0.918	0.978	0.961	to 0.990

Table 4. Sensitivity and Specificity analysis for standalone BC CLA for different cut-off thresholds – BI-RADS 3 and 4 (*for $p < 0.0001$*)

Figure 3 below is an ROC AUC (empirical) curve plot for "standalone" BC CLA computing performance on the validation cohort.

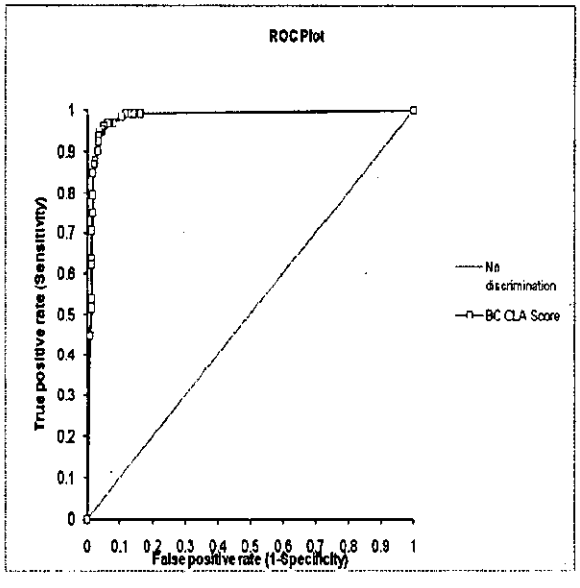

Fig. 3. ROC curve and AUC for "standalone" BC CLA computing performance.

Almen Laboratories, Inc.
1672 Gil Way, Vista CA 92084
(760) 806-0040(voice), (760) 806-7577(fax)
e-mail: info@almenlabs.com http://www.almenlabs.com

For AUC ROC, BC CLA reached 98.2% for empirical and 98.6% for fitted method. At traditional BI-RADS Category 3, which corresponds to BC CLA 3 score, Sensitivity reached 95.5% with Specificity is 94.6% while the most effective BC CLA performance was found at 2.5 threshold level.

BC CLA does not produce the numeric estimation on the same scale as radiologists do when they go through the BI-RADS Assessment process and summarize their assessment in selection of one of the BI-RADS Categories. Therefore to illustrate the comparative performance of BC CLA, Table 5 below summarizes the fitted ROC computations (JROCKFIT software, method developed by professor Metz of Chicago University and used in their clinical trial at John Hopkins) based on the "without" reading data of the validation group of radiologists.

596 cases	*RADs Average*	BC CLA
ROC AUC (A_z fitted)	83.1%	98.7%
Sensitivity	98.7%	98.5%
Specificity	41.5%	89.0%

Table 5. Area (A_z), Sensitivity and Specificity comparison of BC CLA with FROC results of 4 radiologists reading validation data set of 596 cases "without" the device

All the differences with the exception of differences in Sensitivity exceed the statistical variability about the average values. Therefore we can conclude that we have evidence that the "standalone" safety of the device exceeds that of the comparable radiologists' results.

Reproducibility of BC CLA

CADx reproducibility performance was compared for two views of the same mass, radial and anti-radial, acquired during the same examination for 125 masses. An example of such an image pair is shown in Figure 4 for a malignancy. These images show some differences in shape, echo pattern, boundary, and posterior shadowing. One might expect greater variability in malignancies and solid benign masses with perhaps less variability in complicated cysts and simple cysts. The measured CLA values for the image pairs are listed in Table 6 by type of mass.

No significant differences were found. The areas under the ROC curves for the two views were likewise not significantly different (Table 6): 0.95 ± 0.02 for radial and 0.97 ± 0.02 for anti-radial. Table 7 also shows ROC AUC results when the CLA values for each of the image pairs are used separately or when they are combined in various ways.

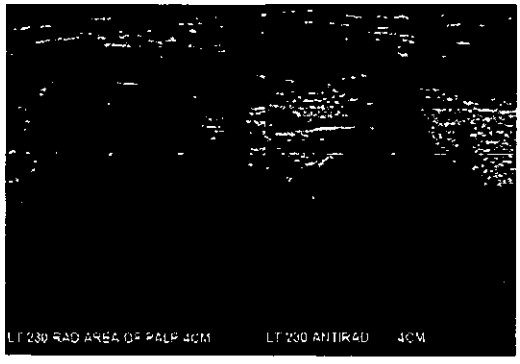

Fig. 4. Radial (left) and anti-radial views of the same mass (malignant).

Lesion Type	Radial Mean ± SD	Anti-Radial Mean ± SD	2-tailed *p* value
Simple Cyst	2.00 ± 0.00	2.03 ± 0.18	0.324*
Complicated Cyst	2.51 ± 0.85	2.36 ± 0.75	0.531*
Solid Benign	2.58 ± 0.92	2.66 ± 0.96	0.599*
Malignant	4.59 ± 0.54	4.67 ± 0.51	0.540*

* *These p-values are not significant*

Table 6. CLA Values for Two Views of Same Mass (125 cases, 30% malignant)

	A_z Fit	± SE	A_z Trapezoidal	± SE
Radial	0.95	0.0207	0.96	0.0275
Anti-Radial	0.97	0.0229	0.97	0.0242

Table 7. ROC AUC for Two Views of the Same Mass (125 cases, 30% malignant)

An additional 36 cases of the same mass scanned on more than one scanner were analyzed. All the differences did not exceed the statistical variability above the average values. Therefore we can confidently conclude that we have clear evidence of proven reproducibility for the device tested performance.

Clinical Validation Data Collection Method

Four sub-specialty breast imaging radiologists independently, in 6 random sessions, assessed the total of 596 US cases (100 randomly selected cases for Sample Size estimation and the balance of 496 cases in validation reading) using a standard BI-RADS hardcopy Classification Form (ACR) that includes final assessment category, descriptors and recommendations for follow-up interval or biopsy. After at least 6 to 9 months' interval from the "without" reading ("washout period"), the same group of radiologists again in 6 random sessions, completed a BI-RADS assessment of the same validation set of randomly mixed cases while using the CADx. In addition to BI-RADS Assessment, the software recorded the radiologist selected Probability of Malignancy. The readers were not aware of the confirmed findings for any case for both reads – "without" and "with" readings.

Clinical Validation Data Analyses Methods

The Clinical Validation Study "with" and "without" the BC device provided effectiveness data via reader performance testing using a Multiple Reader/Multiple Cases (MRMC) design, to answer the question "*Does the physician improve her/his overall assessment performance when using the CAD device?*" The performances of the radiologists (BI-RADS Assessment) "without" and "with" Breast Companion® were compared, to determine if the radiologist improved his/her performance with the use of the CADx software. Primary endpoints were defined as:

- ROC-AUC ("A_z" or Area Under ROC Curve)
- Sensitivity of Breast Companion at BI-RADS cut-point 3, for a 5-scale
- Specificity of Breast Companion at BI-RADS cut-point 3, for a 5-scale

Clinical Validation Reading Results

Summarized ROC analysis results for the validation cohort of 596 cases are presented in Table 8. The AUC differences between the radiologists for "without" reading are not significantly different and are similar to those achieved by both clinical groups in prior studies, and are similar to those reported elsewhere for breast sonography.

	A_z	Sensitivity Cutoff 3	Specificity Cutoff 3	Sensitivity Cutoff 4	Specificity Cutoff 4
Rad1 without CAD	81.88%	97.74%	54.00%	28.57%	98.49%
Rad1 with CAD	85.22%	97.74%	57.02%	39.85%	98.27%
Rad2 without CAD	81.81%	98.50%	41.90%	45.11%	96.54%
Rad2 with CAD	86.76%	96.99%	61.99%	43.61%	96.98%
Rad3 without CAD	83.99%	100.00%	25.92%	87.97%	76.89%
Rad3 with CAD	86.29%	99.25%	34.77%	73.68%	90.28%
Rad4 without CAD	84.80%	98.50%	44.06%	54.89%	96.11%
Rad4 with CAD	90.77%	93.23%	71.71%	66.92%	96.11%
Average "without"	83.12%	98.68%	41.47%	54.14%	92.01%
Average "with"	87.26%	96.80%	56.37%	56.02%	95.41%
Average Difference	**4.14%**	**-1.88%**	**14.90%**	**1.88%**	**3.40%**
Parameters of statistical difference	p=0.05	CI 95% interval size "without" 4.75%	CI 95% interval size "without" 8.97%	CI 95% interval size "without" 15.68%	CI 95% interval size "without" 4.42%

Table 8. Summary of "without" and "with" Validation Study on the set of *596* confirmed cases, cancer prevalence 22.3%.

In addition to effectiveness estimation based on ROC AUC, a traditional accuracy index based on TP, TN, FP and FN differences between "without" and "with" was also computed. Traditional estimation could be computed by using Accuracy index A_c= (TP+TN)/(TP+TN+FP+FN). Using the statistically significant input from the readings ROC results the A_c increase "with" compared to radiologist performance "without" were computed as +6.51% (for 596 case cohort) on average.

Statistical significance of the endpoints

As part of the statistical analyses, the hypothesis H_0 that "without" and "with" treatments are on average equal was tested. Readers were treated as "random" effects in the model because there was no pre-selection or filtration of the readers. With p = 0.0112 for the 596 case cohort, the software confirmed that for ROC AUC, "treatments are not equal" and therefore there is a statistically significant difference between the two "treatments" or "modalities", with the "with" modality being more accurate than the "without" modality.

Statistical Significance of Specificity Improvement "with" CAD is an important co-primary endpoint result. It was found that Specificity improvement of the "with" reading is statistically significant compared to the Specificity of the "without" reading. DBM software output confirms that Specificity of "without" and "with" treatments "are not equal" while for Sensitivity "The treatment SENSITIVITIES were not significantly different" ($p>0.5$).

Safety and Effectiveness evaluation results

The summary results of BC standalone performance and reproducibility evaluation described above and ROC data of BC performance compared to that of the radiologists (Tables 5-7; Fig. 3) support the conclusion that the device's performance is safe.

The clinical Validation Study provided conclusive and statistically significant data (Table 8) for the effectiveness of the BC device and answered the question "*Does the physician improve her/his overall assessment performance when using the CAD device*?" in the affirmative.

Applicability of the Validation Study and statistical results for the device

Tiered effectiveness of CAD evaluation was analyzed. Per FDA recommendations: "The first tier, technical effectiveness, includes measurements of the physical performance of imaging systems, as well as other bench tests and standalone measurements on CADs. The second tier, diagnostic accuracy, includes measurement of sensitivity, specificity, the ROC curve and its summary measures. The first two tiers are more commonly the ones on which industry sponsors of imaging technologies focus when seeking entry to the marketplace through the Food and Drug Administration." Therefore, the described above Validation Study and statistical results demonstrate sufficient evidence of the effectiveness of the device.